



You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice

Title: Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN

Author: Anna Małgorzata Kamińska

Citation style: Kamińska Anna Małgorzata. (2017). Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN. "Toruńskie Studia Bibliologiczne" (2017, nr 2, s. 163-187), doi 10.12775/TSB.2017.022



Uznanie autorstwa - Bez utworów zależnych Polska - Ta licencja zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu zarówno w celach komercyjnych i niekomercyjnych, pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego



ISSN 2080-1807

Anna Małgorzata Kamińska

Instytut Bibliotekoznawstwa i Informacji Naukowej

Uniwersytet Śląski w Katowicach

email: anna.kaminska@us.edu.pl

Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN

DOI: <http://dx.doi.org/10.12775/TSB.2017.022>

STRESZCZENIE: Artykuł przedstawia koncepcję wizualizacji wybranych wskaźników bibliometrycznych na podstawie danych zgromadzonych w bibliograficznej bazie danych CYTBIN, obejmującej artykuły z wybranych krajowych czasopism z zakresu bibliotekoznawstwa i informacji naukowej. Poza zobrazowaniem tradycyjnych wskaźników przedstawiono również wizualizacje miar (np. PageRank, HITS), których wartości nie zależą jedynie od najbliższego sąsiedztwa badanych obiektów w sieci współpracy czy cytowań, ale od szerszego kontekstu wystąpienia w całej strukturze. Celem artykułu nie jest wyciąganie szczegółowych wniosków opisujących rozwój dziedzin bibliotekoznawstwa i informacji naukowej, gdyż wnioski takie – wywiedzione na podstawie danych częściowych – obarczone mogłyby być dużym marginesem niepewności. Zamierzenie autorki stanowiło natomiast przedstawienie samych koncepcji prezentacji danych bibliograficznych w formach, które czyniłyby ich odbiór bardziej zrozumiały i atrakcyjny od tradycyjnych form tabelarycznych.

SŁOWA KLUCZOWE: badania, bibliometria, CYTBIN, wizualizacje, wskaźniki.

Wprowadzenie

Baza CYTBIN to jedna z nielicznych krajowych ogólnodostępnych bibliograficznych baz danych, gromadzących informacje z dziedziny bibliotekoznawstwa, informacji naukowej oraz dyscyplin pokrewnych. Zawiera ona informacje opisujące dziesięć polskich czasopism naukowych w wybranych okresach. Szczegółowe informacje na temat tego projektu, prowadzonego na Uniwersytecie Śląskim, znaleźć można w artykule autorstwa badaczy bezpośrednio zaangażowanych w jej rozwój¹.

Niniejszy artykuł stanowi rozszerzenie badań autorskich opisanych w artykule przedstawiającym potencjał bibliometryczny bazy CYTBIN², gdzie na jej przykładzie (dla stanu na dzień 6 grudnia 2016 r.) przeprowadzono analizy bibliometryczne zgodnie z koncepcjami zaprezentowanymi w autorskiej propozycji wykorzystania miar analiz sieci społecznościowych³. Rozszerzenie to polega w głównej mierze na bogatej wizualizacji i omówieniu uzyskanych tam wyników badań, które trudno byłoby przedstawić w całości w ramach jednego opracowania, głównie z powodu ograniczonej jego objętości i monochromatycznej formy wydawniczej. Podobnie jak w przypadku tamtych badań, tak i tutaj podkreślić należy, że wyniki nie mogą stanowić oparcia dla całościowej oceny poszczególnych naukowców czy rozwoju dziedziny, gdyż dane źródłowe ani nie stanowiły próby reprezentatywnej w rozumieniu badań statystycznych, ani nie były zgromadzone w relacjach sprzyjających ich analizowaniu metodami wychodzącymi poza proste uwzględnianie jedynie najbliższego sąsiedztwa wierzchołków sieci cytowań czy współpracy. Więcej informacji o metodzie pozwalającej gromadzić dane w bardziej sprzyjających ukła-

¹ A. Seweryn, I. Swoboda, *CYTBIN – krajowy indeks cytowań z zakresu bibliologii i informatologii* [online] [dostęp 10 kwietnia 2016]. Dostępny w World Wide Web: <http://eprints.rclis.org/22524/1/cytbin.pdf>.

² A. M. Kamińska, *Potencjał bibliometryczny bibliograficznej bazy danych CYTBIN w świetle prostych i złożonych wskaźników analitycznych*, „Bibliotheca Nostra” 2017 [w druku].

³ A. M. Kamińska, *Zastosowanie metod analizy sieci społecznościowych w dziedzinach bibliometrii i webometrii. Miary i narzędzia*, „Nowa Biblioteka. Usługi, technologie informacyjne i media” 2018, nr 2 [w druku].

dach przedstawiono w opracowaniu dotyczącym metody ProBIT⁴. Tam też, jako przykład tradycyjnej bibliograficznej bazy danych, omówiono logiczny model danych zastosowany przez twórców bazy CYTBIN. Z kolei sposób ekstrakcji danych wykorzystanych do dalszych analiz został przedstawiony w pracy dotyczącej wykrywania powielonych opisów bibliograficznych⁵, operującej również na przykładzie CYTBIN.

Głównymi motywacjami realizacji niniejszych badań była natomiast chęć przybliżenia zainteresowanym bardziej zaawansowanych metod prowadzenia badań bibliometrycznych, wykorzystujących koncepcje zaczerpnięte z aktualnie stosowanych wskaźników oceny parametrycznej naukowców (np. Index Hirscha) i metod analizy sieci społecznościowych oraz przedstawienie możliwości ich wizualizacji wpisujących się w bardzo dynamicznie rozwijającą się obecnie dziedzinę wizualizacji informacji zwaną *InfoVis* (ang. *Information Visualization*). Przykłady zastosowania technik InfoVis w wielu innych dziedzinach znaleźć można chociażby w opracowaniach Veslavy Osińskiej⁶ i Davida McCandlessa⁷.

Współpraca autorów

Pierwszą z badanych zależności stanowiła współpraca pomiędzy badaczami rozumiana jako współautorstwo artykułów naukowych. Il. 1 przedstawia wizualizację współpracy pomiędzy autorami „rodzimymi”, czyli takimi, którzy byli autorami źródłowych artykułów cytujących. Z natury rzeczy są to głównie naukowcy krajowi. Widać wyraźnie, że wybiórczy charakter bazy CYTBIN oraz lokalny (krajowy) charakter dziedziny opisywanej badanymi danymi zaowocował powstaniem kilkusobowych skupisk (klastrów), z których każde dla lepszej identyfikacji oznaczone jest osobnym kolorem. Zastosowany algorytm rozmieszczania, oparty na symulacji sił grawitacji, których oddziaływanie jest tym większe, im

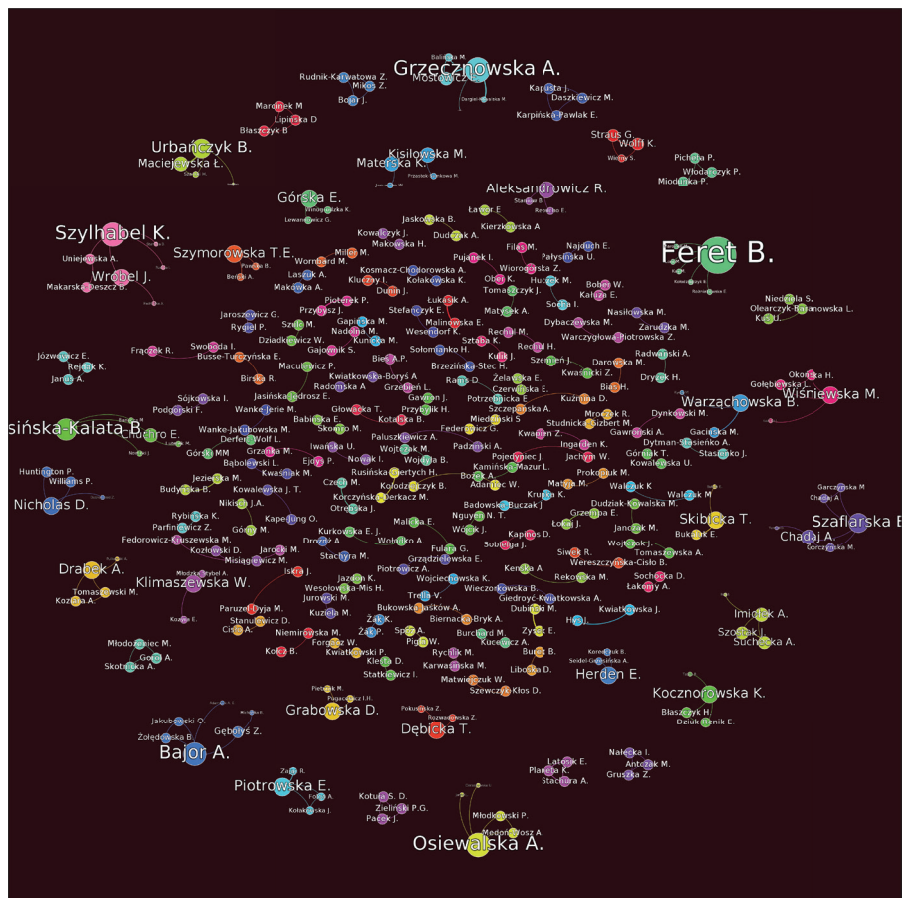
⁴ A. M. Kamińska, *ProBIT – Prospektywna metoda tworzenia indeksów cytowań a problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych*, „Zagadnienia Informacji Naukowej” 2017, nr 1.

⁵ A. M. Kamińska, *Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych*, „Przegląd Biblioteczny” 2017, nr 4 [w druku].

⁶ V. Osińska, *Wizualizacja informacji. Studium informatologiczne*, Toruń 2016.

⁷ D. McCandless, *Informacja jest piękna*, Warszawa 2016.

blizsza jest współpraca danych naukowców, pozwolił na takie rozmieszczenie węzłów o tych samych kolorach (czyli połączonych więzami współpracy) w bliskim wzajemnym sąsiedztwie. Dodatkowo mniej liczne grupy umieszczono w centrum rysunku, natomiast przypadki bardziej licznych grup znalazły swoje miejsce na jego obrzeżach.



Ilustracja 1. Wizualizacja współpracy autorów „rodzimych”

Źródło: opracowanie własne.

Warto zwrócić uwagę, że wielkość wierzchołka wyraża miarę PageRank⁸ dla sieci współpracy⁹ i można ją interpretować w ten sposób, że im większa jest wielkość danego wierzchołka, tym bardziej aktywny był dany naukowiec w kategoriach współpracy z innymi. Wśród największych wierzchołków zaobserwować możemy dwa skrajne przypadki. Jeden to duży wierzchołek połączony z licznymi bardzo małymi punktami niepołączonymi ze sobą (przykładowo Feret Błażej) oraz duży wierzchołek połączony z niemałymi punktami, które również są ze sobą połączone (przykładowo Bajor Agnieszka, Sosińska-Kalata Barbara, Szyhlabeł Krystyna). Ten pierwszy przypadek wskazuje na publikacje dwóch autorów (w przypadku Feret B. wśród źródeł cytujących baza CYTBIN zarejestrowała ich pięć), drugi zaś na publikacje dwu- bądź wieloautorskie (w przypadku Bajor A. są to dwie publikacje dwóch oraz jedna trzech autorów). Na takich podsięciach zaobserwować można również przypadki współpracy niebezpośredniej (na przykład Bajor A. współpracowała z Gębołysiem Zdzisławem, który to z kolei współpracował z Michalską Bożeną zaznaczoną na rysunku bardzo małym punktem).

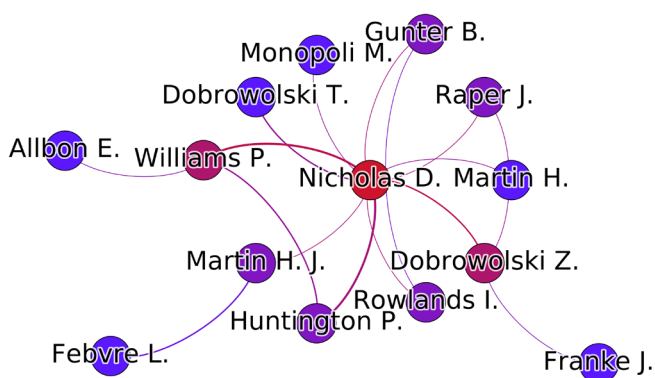
Analizując zjawiska współpracy między badaczami, może zaistnieć potrzeba skupienia się na konkretnych nazwiskach. Warto wtedy sięgnąć po miarę Erdősa (ang. *Erdősnumber*) opisującą „odległość współpracy”. Aby ją otrzymać, trzeba współtworzyć artykuł z kimś, kto miał już ją wcześniej nadaną. Wybrany autor ma wartość $k=0$. Każdy współautor dostaje liczbę $k+1$, gdzie k jest najniższą wartością wszystkich współautorów danego artykułu. Przykładowe wyliczenia tej miary dla wybranych autorów (Nicholas David, Onwuegbuzie AnthonyJ.) przeprowadzono, opierając się na wszystkich jednostkach bibliograficznych zgromadzonych w bazie CYTBIN.

Ilustracja 2 przedstawia wierzchołki z wartościami miary Erdősa 0, 1, 2 dla Nicholas D., który posiadając wartość 0, został umieszczony w samym centrum. Wierzchołki z miarą 1 (oznaczającą bezpośrednią

⁸ A. Langville, C. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton 2012.

⁹ Choć w odniesieniu do przeszukiwania zmiennych w czasie zasobów sieci Internet algorytm PageRank, stanowiący przedmiot patentu, stracił na znaczeniu, to do analizowania struktur sieci społecznościowych (ang. *Social Network Analysis*) używa się współcześnie opartych na nim implementacji wbudowywanych w poszczególne narzędzia analityczne.

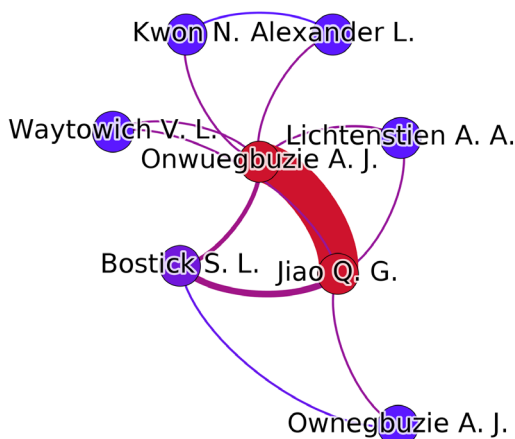
współpracę) zostały rozmieszczone na mniejszej orbicie, a z miarą 2 (oznaczającą pośrednią współpracę) – na orbicie zewnętrznej. Łuki łączące poszczególne wierzchołki obrazują współpracę pomiędzy poszczególnymi autorami i nie ograniczają się one jedynie do tych wychodzących od Nicholas D. Warto zwrócić uwagę na kolory wierzchołków, którymi oznaczono nieważony stopień (czyli analizowano różnorodność współpracy, a nie siłę poszczególnych związków) – im cieplejszy kolor, tym wyższy stopień wierzchołka (wartość stopnia wierzchołka dla Nicholas D. wynosi 10). Należy zaznaczyć, że prezentowany barwą stopień wierzchołka został wyliczony dla całości bazy danych, tak więc nic nie stałoby na przeszkodzie, gdyby prezentowany podgraf zawężony dla liczby Erdősa w odniesieniu do Nicholas D. mniejszej lub równej dwa posiadał na swoich obrzeżach wierzchołki w kolorze cieplejszym od wierzchołka centralnego.



Ilustracja 2. Miara Erdősa 0,1,2 dla Nicholas D.

Źródło: opracowanie własne.

Ilustracja 3 przedstawia natomiast podgraf dla liczby Erdősa w odniesieniu do Onwuegbuzie A. J. o wartościach 0, 1, 2. Warto tutaj zwrócić uwagę na grubości łuków łączących poszczególnych autorów obrazujące siły łączących ich więzi. Jak wynika z rysunku, badany autor najsilniej współpracował z Jiao Qun G. (kilkanaście wspólnych artykułów). Temperatura barwy wierzchołka wyraża tutaj stopień ważony (a więc uwzględnia siły poszczególnych związków) i dla Onwuegbuzie A. J. wynosi ona 24.



Ilustracja 3. Miara *Erdősa* 0, 1, 2 dla Onwuegbuzie A. J.

Źródło: opracowanie własne.

Interesujące spostrzeżenie poczynić można, obserwując orbitę zewnętrzną, na której znaleźć można autora Ownegbuzie A. J. Z dużym prawdopodobieństwem (świadczy o tym zbieżność nazwisk oraz fakt współpracy z Jiao Q. G.) powiedzieć można, że jest to tak naprawdę badany autor z błędem literowym popełnionym przez operatora bazy danych podczas wprowadzania rekordu opisującego jednostkę przedstawioną na il. 4.

CYTOWANIE - WYDAWNICTWO ZWARTE	
Autorzy:	Bostick S. L., Ownegbuzie A. J., Jiao Q. G.
Rodzaj nośnika:	D
Tytuł:	Library anxiety. Theory, research and applications
Miejsce wydania:	Lanham, Md
Rok wydania:	2004

Ilustracja 4. Dane szczegółowe rekordu bibliograficznego dla Ownegbuzie A. J.

Źródło: CYTBIN [online] [dostęp 6 grudnia 2016]. Dostępny w World Wide Web: http://www1.bg.us.edu.pl/bazy/cytbins/cytowania_wz.asp?id=1962.

O metodach wspierających operatorów bibliograficznych baz danych w niepopelnianiu tego typu błędów przeczytać można we wspomnianym wcześniej opracowaniu dotyczącym metody ProBIT¹⁰, natomiast o metodach umożliwiających automatyczne wspieranie w wykrywaniu takich błędów – chociażby w artykule opisującym metody deduplikacji rekordów bibliograficznych¹¹.

Cytowania autorów

Kolejną z analizowanych relacji łączących poszczególnych badaczy jest współpraca rozumiana jako korzystanie z dorobku innych celem doskonalenia go, czyli wynikająca z artykułów cytowanych, zamieszczanych w bibliografii załącznikowej jednostek cytujących.

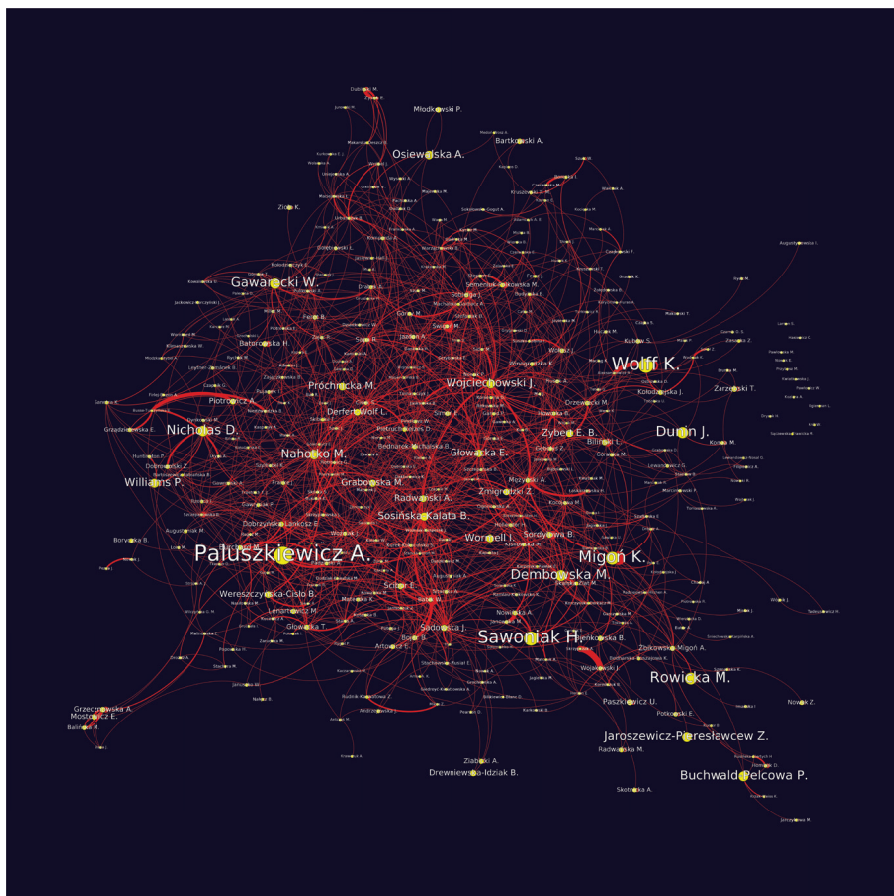
Cytowania prac naukowych stanowią jedną z najważniejszych perspektyw zarówno parametrycznej oceny wkładu dorobku konkretnego badacza w rozwój danej dziedziny nauki, jak również oceny i kierunków rozwoju dziedzin nauki obserwowanych globalnie.

Ilustracja 5 przedstawia sieć cytowań pomiędzy autorami ograniczoną (dla przejrzystości formy) do autorów o „cytowalności” prezentującej arbitralnie przyjęty próg. Należy zwrócić uwagę, że w przeciwieństwie do relacji współautorstwa, cytowanie nie jest relacją zwrotną, tak więc przedstawiony graf jest grafem skierowanym. Kierunek relacji w przyjętej na rysunku wizualizacji zgodny jest z kierunkiem ruchu wskazówek zegara. Zastosowana technika wizualizacji, podobnie jak w przypadku kilku kolejnych prezentowanych tutaj obrazów, uzależnia rozmiar wierzchołka i jego etykiety od wartości wyliczonej miary. Pozwala to zwrócić uwagę na najważniejsze względem danej miary obiekty na tle mniej istotnych, które stanowią jedynie tło i kontekst wizualizacji. Podobnie jak tradycyjnie w formie tabelarycznej prezentuje się zwykle – jako wyniki badań – jedynie ułamek najważniejszych względem danej miary wierszy, a dalsze szczegółowe badania prowadzić można np. za pomocą arkuszy kalkulacyjnych, które pozwalają na „przewijanie danych”, tak w przy-

¹⁰ A.M. Kamińska, *ProBIT – Prospektywna metoda tworzenia trawersowalnych indeksów cytowań...*

¹¹ A. M. Kamińska, *Miary podobieństw łańcuchów znakowych...*

padku analiz struktur sieciowych dedykowane im narzędzia pozwalają na interaktywne przybliżanie pewnych regionów grafów.



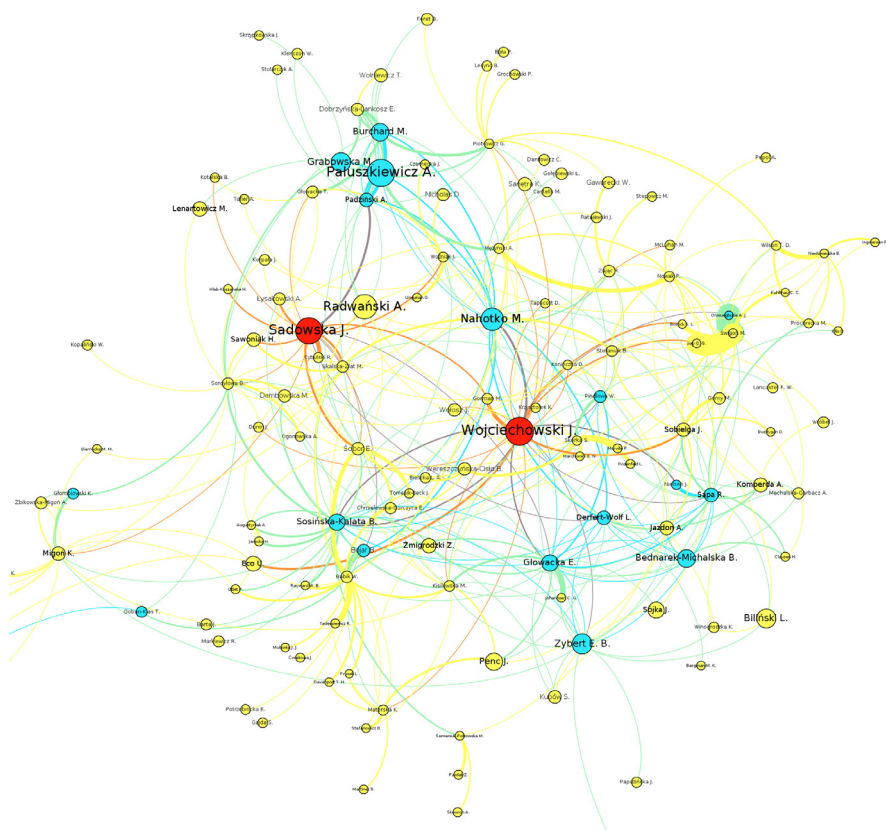
Ilustracja 5. Ograniczone cytowania autorów

Źródło: opracowanie własne.

Wielkość wierzchołków wynika bezpośrednio z miary PageRank wyliczonej dla całości sieci cytowań. Zaobserwować można przy jej pomocy autorów cytowanych najczęściej, a na czoło wysuwają się naukowcy tacy, jak Paluszkiwicz Anna, Wolff Katarzyna, Sawoniak Henryk, Migoń Krzysztof oraz Rowicka Małgorzata. Podobnie jak w przypadku il. 1,

tak i tutaj wykorzystano algorytm rozmieszczania uwzględniający siłę powiązań pomiędzy poszczególnymi wierzchołkami, tak więc odległość i rozmieszczenie poszczególnych wierzchołków nie są przypadkowe, a wynikają bezpośrednio ze ścieżek cytowań. Im bliższe sobie są poszczególne wierzchołki, tym ściślej współpracują (w sensie pogłębiania dorobku naukowego autora cytowanego) ze sobą poszczególni badacze.

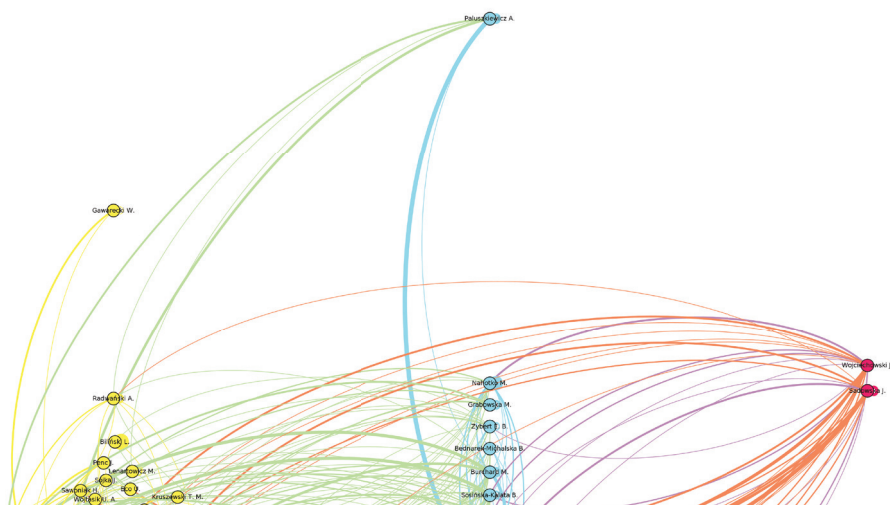
Wizualizacja przedstawiona na il. 6 stanowi niejako rozszerzenie wcześniejszej prezentacji o zobrazowanie wskaźnika Hirscha i usunięcie węzłów ze wskaźnikiem niższym niż 2.



Ilustracja 6. Mapa cytowań z PageRank obliczonym bez autocytowań oraz wskaźnikiem Hirscha

Źródło: opracowanie własne.

Rozmiar wierzchołka w dalszym ciągu stanowi przedstawienie miary PageRank, wskaźnik Hirscha obrazowany jest barwą (żółta – 2, turkusowa – 3, czerwona – 4). Warto zauważyć, że niektóre wierzchołki o podobnych rozmiarach mają różne wskaźniki Hirscha. Spostrzeżenie to zachęciło autorkę do dalszych badań, których wizualizację przedstawiono na il. 7.

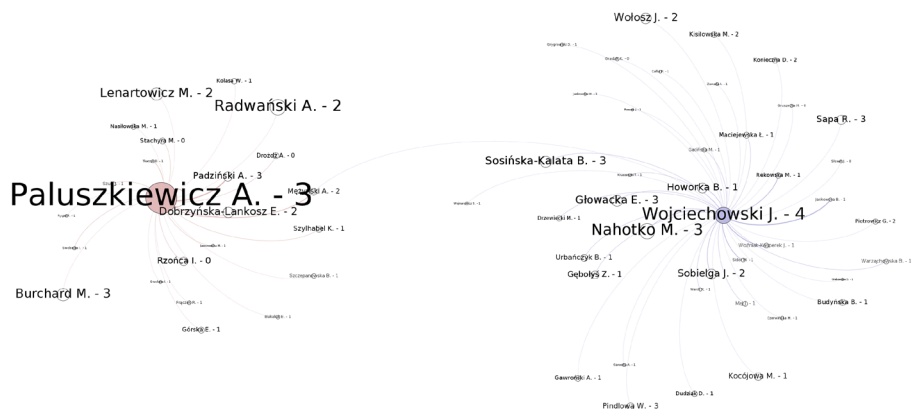


Ilustracja 7. Przedstawienie zależności miary PageRank od wskaźnika Hirscha (wycinek)

Źródło: opracowanie własne.

Mapę przedstawioną na rysunku należy interpretować w ten sposób, że położenie na osi odciętych wyraża wartość wskaźnika Hirscha, który podobnie jak na poprzednim rysunku zobrażony jest również barwą. Warto zwrócić uwagę, że wartości tego wskaźnika są całkowitoliczbowe, a położenie na osi odciętych na to nie wskazuje. Stało się tak dlatego, że z powodu dużego zagęszczenia żółtych wierzchołków zastosowano dodatkowe rozmieszczenie poprawiające czytelność poszczególnych nazwisk. Położenie na osi rzędnych wyraża natomiast wartość miary PageRank. Na przedstawionym rysunku zaobserwować możemy niepokojący brak „względnej monotoniczności” obydwu miar. Innymi słowy, można by się

jednak różne wartości wskaźnika Hirscha – liczby umieszczone po nazwiskach wskazują wartość wskaźnika Hirscha.



Ilustracja 9. Najbliższe otoczenie dwóch liderów miary PageRank

Źródło: opracowanie własne.

Widzimy, że badaczy tych łączy Mężyński Andrzej, który jako jedyny cytował prace jednego i drugiego. Ważniejsze spostrzeżenie opierać się można na porównaniu cytowania przez innych badaczy z wysokimi wartościami wskaźnika Hirscha. Otóż Paluszkiwicz A. cytowana była dwukrotnie przez innych autorów ze wskaźnikiem Hirscha o wartości 3 (Burchard Maria, Padziński Andrzej), natomiast Wojciechowski J. miał takich cytowań aż 3 (Sosińska-Kalata B., Sapa Remigiusz, Pindłowa Wanda). Zauważyć możemy również, że ogólna liczba cytowań w przypadku Wojciechowskiego J. jest wyższa, chociaż pochodzi ona od autorów z mniejszymi wierzchołkami. Ponieważ miara PageRank uwzględnia także dalsze sąsiedztwo wierzchołków, to przeanalizować by należało również sąsiedztwo sąsiedztwa badanych autorów, co w sposób rekurencyjny zmusiłoby nas do przeanalizowania większej części sieci.

Abstrahując od rozsądzania, która z miar w sposób bardziej intuicyjny oddaje wartość wkładu badacza w rozwój dziedziny, zauważyć należy, że do oceny parametrycznej mającej na celu wartościowanie najnowszego dorobku danego badacza miara PageRank nie bardzo się nadaje,

gdyż jest ona zależna również od dalszego jego otoczenia, co czyni ją mocno zmienną w czasie i niemożliwą do zastosowania *ad hoc*.

Pozostając przy analizowaniu relacji cytowań pomiędzy autorami, ale odwracając jej kierunek, przyjrzeć się można autorom cytującym najczęściej. Il. 10 przedstawia wierzchołki o losowo przydzielonych kolorach i wychodzących z nich krawędzie w kolorach danego wierzchołka.

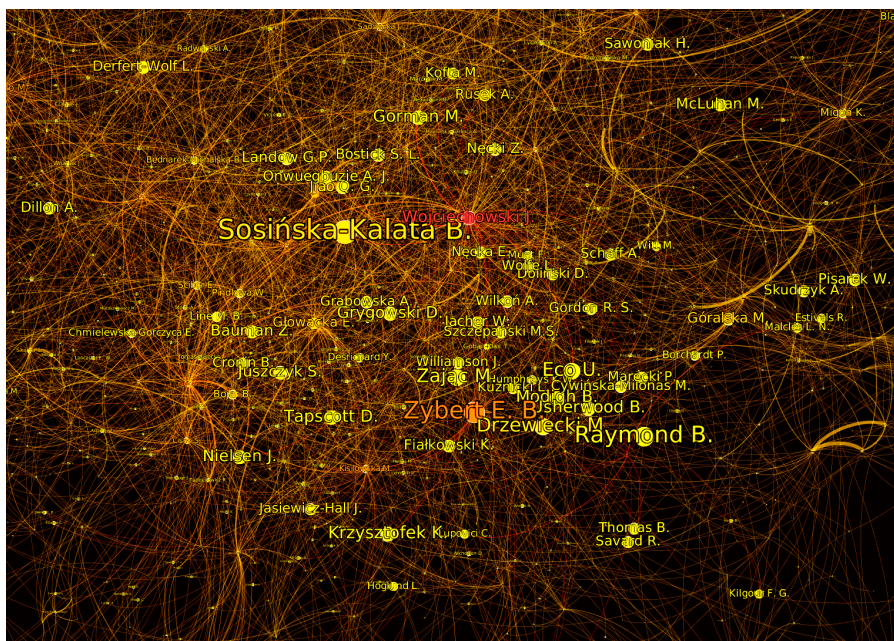


Ilustracja 10. Najaktywniejsi cytujący autorzy

Źródło: opracowanie własne.

Tworzy to swojego rodzaju „warkocze” cytowań, których centra wskazują na autora cytującego. Wielkość czcionki przedstawiającej nazwisko autora wynika wprost ze stopnia wychodzącego wierzchołka (liczby wychodzących z niego krawędzi cytowań), tak więc największe napisy reprezentują najczęściej cytujących autorów.

Analizując ścieżki cytowań pomiędzy autorami, warto przyjrzeć się dwuwartościowemu wskaźnikowi HITS12 zobrazowanemu na il. 11.



Ilustracja 11. Wizualizacja miary HITS dla mapy cytowań

Źródło: opracowanie własne.

Wskaźnik ten zakłada wyliczenie dwóch wartości dla każdego z węzłów – wartości autorytetu (ang. *authority*) oraz koncentratora (ang. *hub*), z których pierwsza wyrażać ma aktywność danego wierzchołka rozumianą jako

¹² S. Cheng, P. YunTao, Y. JunPeng, G. Hong, Y. ZhengLu, H. ZhiYu, PageRank, HITS and Impact Factor for Journal Ranking, [w:] 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles 2009, s. 285–290.

ważność w generowaniu nowych informacji (wkładu w rozwój nauki), a druga aktywność w propagowaniu informacji (upowszechnienie nauki). Wielkość danego wierzchołka wyraża wartość autorytetu, natomiast temperatura barwy wartość koncentratora. Już pierwszym rzutem oka na rysunek pozwala zaobserwować duże wierzchołki o stosunkowo zimnych barwach (np. Sosińska-Kalata B., Raymond Boris), stosunkowo duże wierzchołki o barwach cieplejszych (Zybert Elżbieta Barbara) oraz również duży wierzchołek o barwie najcieplejszej (Wojciechowski J.). Dziwić trochę może stosunkowo nieduży rozmiar wierzchołka reprezentującego Wojciechowskiego J. wobec zarówno posiadania wysokiej wartości miary PageRank, jak również wysokiej wartości wskaźnika Hirscha. Spostrzeżenie to będzie źródłem dociekań dalszych badań autorskich.

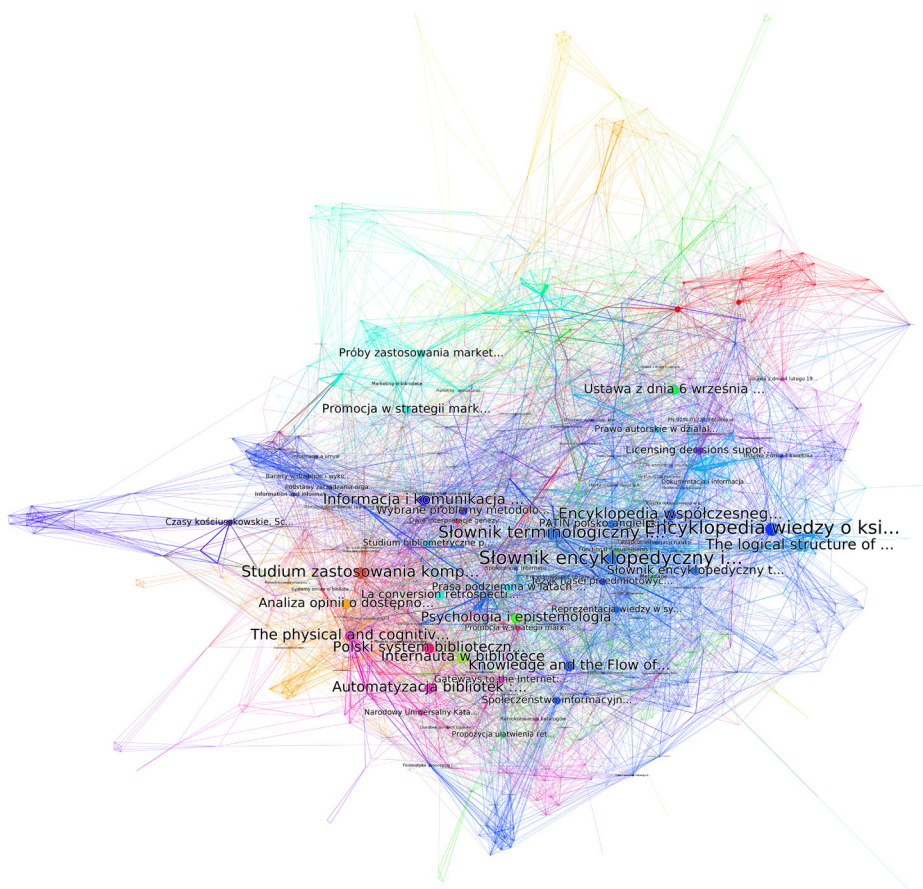
Podobieństwo dokumentów

Zmieniając perspektywę analiz z sieci relacji wiążących autorów na relacje zachodzące pomiędzy poszczególnymi dokumentami, warto się bliżej przyjrzeć metodzie współcytowań¹³ zobrazowanej na il. 12.

Ilustracja 12 przedstawia mapę wszystkich dokumentów, których siła wzajemnych relacji wynika z częstotliwości ich cytowania jednocześnie przez inne dokumenty. Warto zauważyć, że metoda ta daje możliwość analizowania dokumentów cytowanych (w przeciwieństwie do metody powiązań bibliograficznych, która pozwala analizować jedynie dokumenty cytujące), co z jednej strony znacznie rozszerza wolumen analizowanych informacji (zakładając oczywiście, że przeciętnie jeden dokument cytujący przywołuje wiele innych dokumentów), a z drugiej przenosi obszar analiz poza ten, który jest wytyczony zbiorem dokumentów cytujących. Warto również uzmysłowić sobie, że w niektórych przypadkach metoda ta może dawać nie do końca oczekiwane wyniki. Na przykład artykuły przekrojowe czy przeglądowe „generować” będą podobieństwo dokumentów pozornie ze sobą niezwiązanych (artykuł o analizowaniu podobieństwa

¹³ Więcej o metodach współcytowań oraz zobrazowanej metodzie powiązań bibliograficznych zob. np. I. Marszakowa-Szajkiewicz, *Badania ilościowe nauki podejście bibliometryczne i webometryczne*, Poznań 2009.

treści dokumentów wiązał będzie dokumenty z dziedziny lingwistyki oraz algorytmów klastrowania).



Ilustracja 12. Mapa współcytowań jednostek bibliograficznych

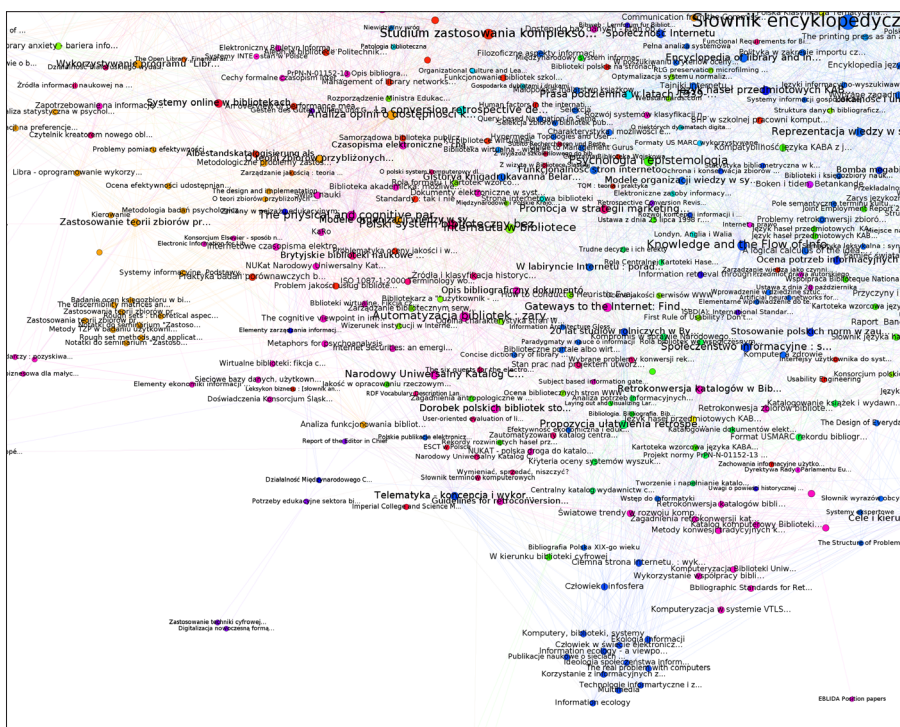
Źródło: opracowanie własne.

Ilustracja 12 przedstawia zgrubny szkic całości powiązań z wybarwieniem automatycznie zidentyfikowanych grup tematycznych oraz przedstawieniem tytułów najważniejszych dokumentów reprezentujących te grupy.

Sposób analizy szczegółowej wybranego fragmentu mapy przedstawia il. 13. W dolnej jej części zaobserwować można niebieskie skupisko

wierzchołków reprezentujące grupę artykułów związaną z zastosowaniem technologii komputerowych do systemów bibliotecznych i wyszukiwawczych, widoczne zaś po lewej stronie skupisko żółtych wierzchołków przedstawia grupę artykułów dotyczącą zastosowania metod ilościowych do danych bibliograficznych. Najłatwiej tego typu analizy prowadzić nie na statycznych obrazach, a na interakcyjnych narzędziach z możliwością przeszukiwania i przybliżania poszczególnych obszarów mapy.

Górna część całości mapy, zawierająca żółte skupisko wierzchołków, została przedstawiona na il. 14.

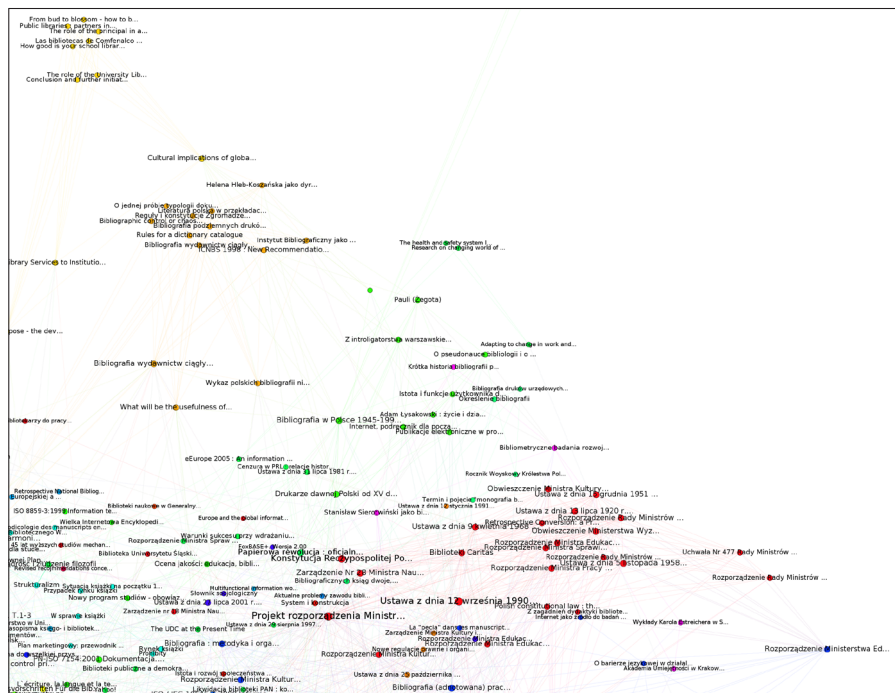


Ilustracja 13. Szczegółowy widok fragmentu mapy współcytowań – dolna część mapy

Źródło: opracowanie własne.

Na tak przybliżonym obszarze mapy widać wyraźniej skupisko czerwonych wierzchołków dokumentów, które łączy tematyka legislacyjna, oraz

w lewym górnym rogu jasnożółte skupisko obcojęzycznych dokumentów dotyczące zapewne funkcjonowania bibliotek zagranicznych.



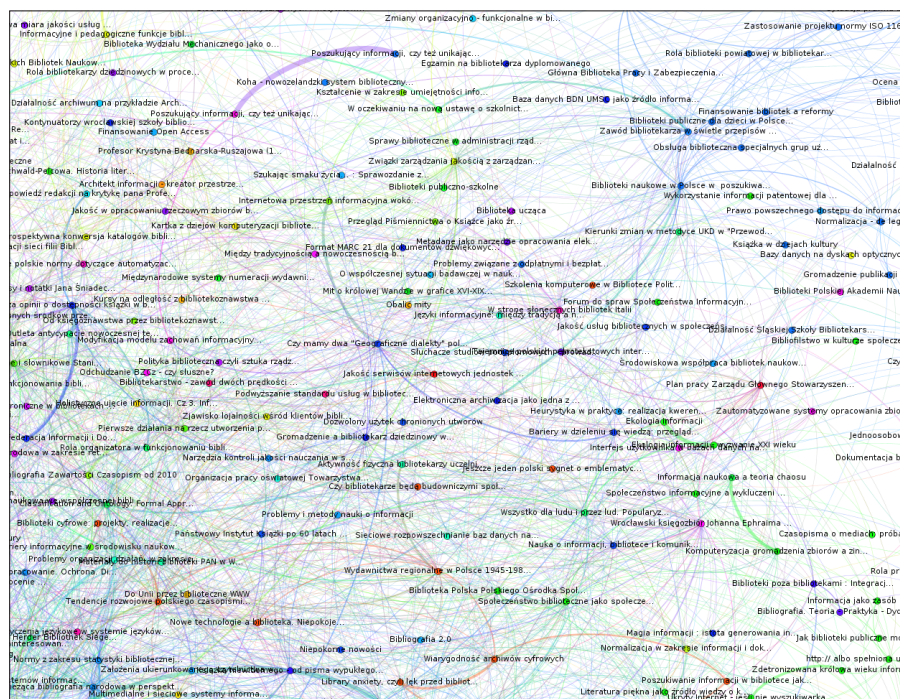
Ilustracja 14. Szczegółowy widok fragmentu mapy współcytowań – górna część mapy

Źródło: opracowanie własne.

Druga z użytych metod badania zależności między dokumentami – metoda powiązań bibliograficznych – pozwala ocenić podobieństwo dokumentów cytujących na podstawie ich bibliografii załącznikowych. Zakłada się, że dwa dokumenty są do siebie tym bardziej podobne, im więcej zidentyfikowano wspólnych jednostek w ich spisach bibliograficznych.

Warto tutaj zwrócić uwagę, że metoda ta stanowi niejako odwrócenie koncepcji poprzedniej i że na podstawie dokumentów cytowanych analizuje się dokumenty cytujące. Wydaje się, że metoda ta ma większe szanse grupowania zgodnego z oczekiwaniami tak kłopotliwych artykułów przeglądowych, gdyż te w sposób naturalny powinny w swojej

bibliografii zawierać po kilka pozycji dotyczących każdego z poruszanych zagadnień. W przywołanym już przykładzie automatycznego grupowania treści tekstów w bibliografii załącznikowej znaleźć się powinny zarówno artykuły dotyczące językoznawstwa, jak i algorytmów automatycznego grupowania. Jeśli część tytułów pojawi się w innych dokumentach, świadczyć to będzie zapewne, że dotyczą one i zagadnień językoznawstwa, i grupowania, a więc prawdopodobnie grupowania tekstów języka polskiego. Przykład szczegółowego fragmentu mapy powiązań bibliograficznych przedstawiono na il. 15.



Ilustracja 15. Szczegółowy widok fragmentu mapy powiązań bibliograficznych
Źródło: opracowanie własne.

Widzimy tutaj poszczególne wierzchołki połączone tym grubszymi łukami, im więcej miały wspólnych pozycji bibliograficznych. Poza próbami interpretacji merytorycznej takiego, a nie innego wzajemnego rozłożenia wierzchołków reprezentujących poszczególne dokumenty, warto zwrócić

uwagę na górną część fragmentu mapy i umiejscowiony tam najgrubszy łuk łączący dokumenty o tych samych nazwach – „Poszukujący informacji, czy też unikający tego zadania użytkownik”. W bazie CYTBIN artykuł ten został zarejestrowany podwójnie, o czym świadczy wynik wyszukiwania w systemie przedstawiony na il. 16. Pomimo zgodności nazwy czasopisma, tytułu i numeru zeszytu warto zwrócić uwagę na rozbieżności dotyczące inicjału imienia (a konkretnie kończącej go kropki), stron oraz liczby cytowań. Jednak nawet mimo dość dużej rozbieżności w cytowanych pozycjach, wspólnie cytowanych jednostek bibliograficznych było na tyle dużo, że artykuł ten dało się łatwo zidentyfikować.

Wybrane kryteria: tytuł=poszukujący informacji	
1.	Niedźwiedzka B: Poszukujący informacji, czy też unikający tego zadania użytkownik PTINT. Praktyka i Teoria Informacji Naukowej i Technicznej 2001 nr 1-2 s.24-30, liczba cytowań: 26
2.	Niedźwiedzka B.: Poszukujący informacji, czy też unikający tego zadania użytkownik PTINT. Praktyka i Teoria Informacji Naukowej i Technicznej 2001 nr 1-2 s.24-31, liczba cytowań: 23

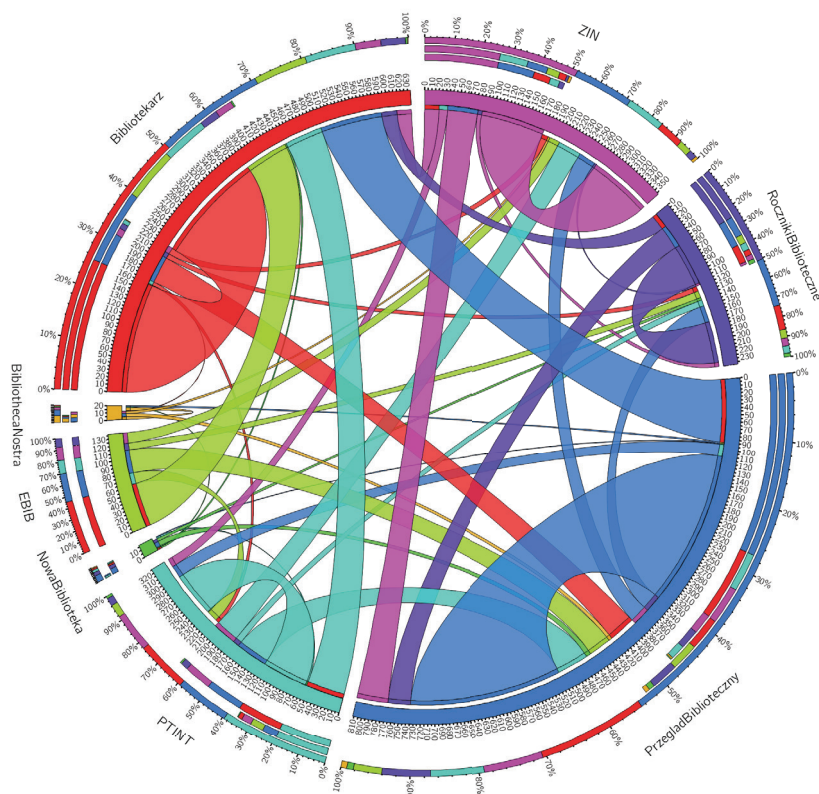
Ilustracja 16. Wynik wyszukiwania w bazie CYTBIN artykułów o tytule *Poszukujący informacji...*

Źródło: CYTBIN [online] [dostęp 6 grudnia 2016]. Dostępny w World Wide Web: http://www1.bg.us.edu.pl/bazy/cytbin/szukaj_cytowania.asp.

Cytowania czasopism

Ilustracja 17 prezentuje zagregowane informacje o cytowaniach pomiędzy czasopismami.

Pokazano na nim zależności pomiędzy wszystkimi czasopismami, dla których w bibliograficznych opisach artykułów gromadzi się informacje o ich bibliografiach załącznikowych. Widzimy tutaj, że do najczęściej cytowanych czasopism zaliczyć należy: „Przegląd Biblioteczny”, „Bibliotekarza”, „Zagadnienia Informacji Naukowej”, „Praktykę i Teorię Informacji Naukowej i Technicznej” oraz „Roczniki Biblioteczne”. Obserwacja skumulowanych wartości fragmentów pierścieni opisujących cytowania (odpowiednio dla kierunków „cytujący” oraz „cytowany”) pozwala wy-



Ilustracja 17. Zagregowany widok cytowań pomiędzy czasopismami

Źródło: opracowanie własne.

snuć wnioski, że najczęściej cytowanym w bazie CYTBIN czasopismem jest „Przegląd Biblioteczny”, który jednocześnie ma największą liczbę autocytowań. Najczęściej cytowanym czasopismem przez inne czasopisma (a więc bez uwzględnienia autocytowań) jest z kolei „Bibliotekarz”, który ma jednocześnie największy stosunek liczby artykułów, które go cytują, do liczby jego artykułów cytujących. Proporcje te są bardziej zrównoważone w przypadku pozostałych wyżej wymienionych czasopism z wyjątkiem „Praktyki i Teorii Informacji Naukowej i Technicznej”, gdzie z kolei liczba artykułów cytujących jest blisko dwukrotnie wyższa niż cytowanych. Przyglądając się strumieniom przepływu wiedzy, to pomi-

jając łuki reprezentujące autocytowania, warto zauważyć, że największe szerokości łuków dotyczą koloru niebieskiego („Przegląd Biblioteczny”), seledynowego („EBIB”) oraz miętowego („Praktyka i Teoria Informacji Naukowej i Technicznej”) i wszystkie one są cytowaniami czasopisma „Bibliotekarz”. Dodatkowo zaobserwować można, że w przypadku czasopism cytowanych najczęściej, autocytowania stanowią dla nich największe źródło przepływu informacji, wynosząc nierzadko około połowę sumy wszystkich artykułów je cytujących.

Trzeba tu jednak raz jeszcze podkreślić, że na dzień dzisiejszy baza CYTBIN nie gromadzi informacji o wszystkich publikacjach omawianych czasopism i dlatego uzyskane wyniki mogą nie oddawać w pełni obrazu faktycznego.

Podsumowanie

W artykule przedstawiono możliwości wizualizacji wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN w postaci sieci współpracy pomiędzy naukowcami, sieci cytowań pomiędzy naukowcami oraz map dokumentów przedstawiających ich powiązania wynikające z metod współcytowań oraz powiązań bibliograficznych. Niektóre z wykorzystanych widoków uzyskano, stosując algorytmy rozmieszczania operujące bezpośrednio na wyliczonych siłach interakcji pomiędzy wierzchołkami. Widoki takie pozwalają na ich bezpośrednią i intuicyjną interpretację, gdyż wierzchołki położone blisko siebie są również sobie bliskie w kontekście danej miary. Inne widoki, umożliwiające rozmieszczanie wierzchołków zgodnie z ich atrybutami odgrywającymi w danej chwili rolę wartości współrzędnych kartezjańskich, wykorzystano do porównania niektórych miar.

Prezentowano sposoby wizualizacji powszechnie już używanych w badaniach bibliometrycznych miar, takich jak liczba cytowań czy wskaźnik Hirscha, oraz takich, które z różnych przyczyn nie znalazły jeszcze w tej dziedzinie powszechnego zastosowania. Analizując niektóre ze zjawisk opisywanych wykorzystanymi miarami, autorka natknęła się na sporadyczne nieprawidłowości w danych zgromadzonych w bazie CYTBIN, co nie zmienia faktu, że dane te są cenne, a inicjatywa ich gromadzenia i uaktualniania powinna być rozwijana i kontynuowana.



Autorka ma również nadzieję, że zaprezentowane wizualizacje, będące częścią szerszej dziedziny wizualizacji informacji, pozwoliły wykazać, że wyjście z warsztatem badawczym poza najczęściej obecnie stosowane układy tabelaryczne i wykresy pozwala dostrzec zjawiska trudne lub wręcz niemożliwe do zidentyfikowania przy wykorzystaniu metod tradycyjnych, a niniejszy artykuł będzie stanowił bodziec do dalszych dociekań dla niej samej oraz innych badaczy.

Bibliografia

- Cheng Su, YunTao Pan, JunPeng Yuan, Hong Guo, ZhengLu Yu, ZhiYu Hu, *Page-Rank, HITS and Impact Factor for Journal Ranking*, [w:] 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles 2009, s. 285–290.
- Kamińska Anna Małgorzata, *Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych*, „Przegląd Biblioteczny” 2017, nr 4 [w druku].
- Kamińska Anna Małgorzata, *Potencjał bibliometryczny bibliograficznej bazy danych CYTBIN w świetle prostych i złożonych wskaźników analitycznych*, „Bibliotheca Nostra” 2017 [w druku].
- Kamińska Anna Małgorzata, *ProBIT – prospektywna metoda tworzenia trawersalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych*, „Zagadnienia Informacji Naukowej” 2017, nr 1.
- Kamińska Anna Małgorzata, *Zastosowanie metod analizy sieci społecznościowych w dziedzinach bibliometrii i webometrii. Miary i narzędzia*, „Nowa Biblioteka. Usługi, technologie informacyjne i media” 2018, nr 2 [w druku].
- Langville Amy, Meyer Carl, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton 2012.
- Marszakowa-Szajkiewicz Irena, *Badania ilościowe nauki podejście bibliometryczne i webometryczne*, Poznań 2009.
- McCandless David, *Informacja jest piękna*, Warszawa 2016.
- OsińskaVeslava, *Wizualizacja informacji. Studium informatologiczne*, Toruń 2016.
- Seweryn Anna, Swoboda Izabela, *CYTBIN – krajowy indeks cytowań z zakresu bibliologii i informatologii* [online] [dostęp 30 września 2017]. Dostępny w World Wide Web: <http://eprints.rclis.org/22524/1/cytbins.pdf>.

Visualizations of Selected Bibliometric Indicators on the Example of the Bibliographic Database CYTBIN

ABSTRACT: The article presents the concept of visualizations of selected bibliometric indicators based on data collected in the CYTBIN bibliographic database that covers articles from selected national journals of information and library science. In addition to traditional bibliometric indicators, visualizations of measures that do not depend only on the nearest neighborhood of the analyzed objects in the citation and collaboration networks, but from a broader context across the structure, are depicted (such as PageRank, HITS). The aim of this article is not to draw detailed conclusions about the development of information and library science, as such conclusions based on incomplete data could be burdened with some error. The intention of the author, however, was to present concepts of visualizations of bibliographic data in forms which make them more readable and attractive than traditional tabular forms.

KEYWORDS: analysis, bibliometrics, CYTBIN, indicators, visualizations.

